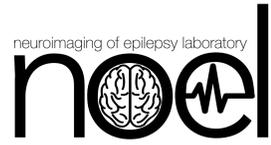


UNCERTAINTY-INFORMED DETECTION OF EPILEPTOGENIC BRAIN MALFORMATIONS USING BAYESIAN NEURAL NETWORKS



Gill RS*, Caldairou B, Bernasconi N, Bernasconi A



Neuroimaging of Epilepsy Laboratory, McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montréal, QC, Canada

PURPOSE

Focal cortical dysplasia (FCD) is a prevalent surgically-amenable epileptogenic malformation of cortical development. On MRI, FCD typically presents with cortical thickening, hyperintensity, and blurring of the gray-white matter interface¹. These changes may be visible to the naked eye, or subtle and be easily overlooked. Despite advances in MRI analytics, current machine learning algorithms²⁻⁵ fail to detect FCD in up to 50% of cases⁶. Moreover, the deterministic nature of current algorithms does not allow conducting risk assessments of such predictions, an essential step in clinical decision-making. Here, we propose an algorithm formulated on Bayesian convolutional neural networks (CNN)⁷ providing information on prediction uncertainty, while leveraging this information to improve classification performance. Our classifier was trained on a patch-based augmented dataset derived from 56 patients with histologically-validated FCD to distinguish the lesion from healthy tissue. The algorithm was trained and cross-validated on multimodal 3 Tesla MRI data. Compared to a non-Bayesian learner with the same network architecture and complexity, the uncertainty-informed Bayesian CNN classifiers showed significant improvement in sensitivity (89% vs 82%; $p < 0.05$) while specificity was high for both classifiers. We demonstrate empirically the effectiveness of our uncertainty-informed CNN algorithm, making it ideal for large-scale clinical diagnostics of FCD.

METHOD

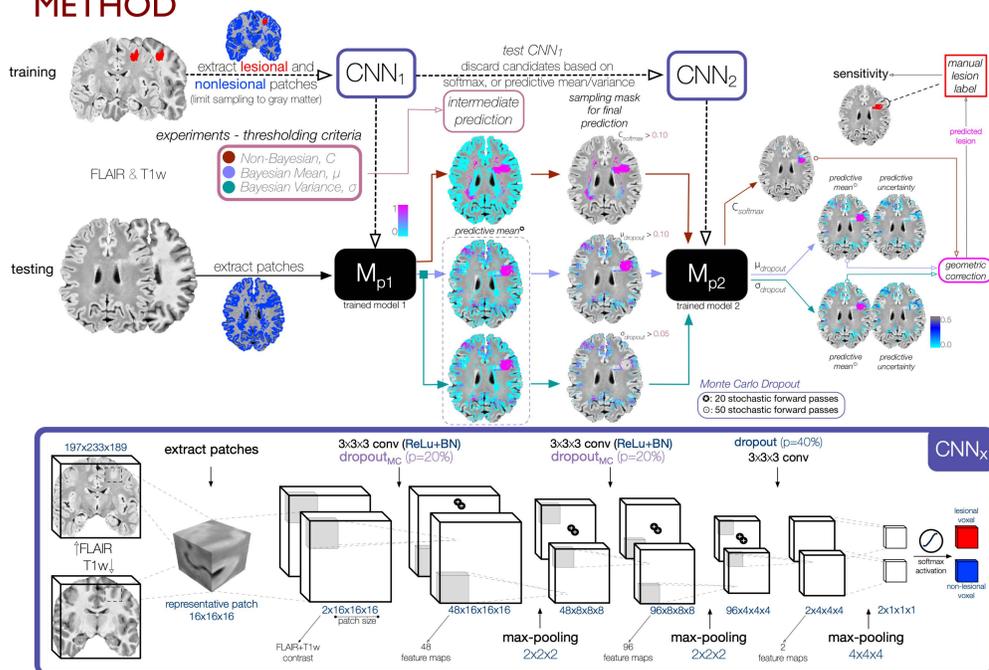


Figure 1. Upper panel: Training and testing schema using two-stage CNN_x cascade (CNN1/CNN2). To reduce training times, multimodal patches (2x16x16x16; centered around the voxel to classify) are sampled from a gray matter (GM) mask. This mask is generated using intra-subject z-score of FLAIR contrast, discarding hypointense voxels ($z < 0.1$). Performance (sensitivity) is measured relative to the expert manual labels. The schema also illustrates the processing workflow of three comparisons [Bayesian (mean- and uncertainty-based) vs Non-Bayesian] undertaken.

Lower panel: Convolutional network architecture (CNN_x) for two-label classification with three consecutive convolutions and max-pooling units, followed by a voxel-wise softmax classification using multimodal (FLAIR+T1w) patches. Each convolution is followed by rectified linear units (ReLU) to introduce non-linearity. Batch normalization (BN) and dropout serve as regularizers. Dropout_{MC} induces stochasticity in the network predictions. Adadelta serves as the optimizer to minimize the binary crossentropy loss.

Training. Volumetric 3T T1-weighted 3D-MPRAGE (T1w) and 3D-FLAIR MRI were collected in 56 patients (mean age: 26 ± 10) with histologically-verified FCD. Routine MRI was initially reported as unremarkable in 80%. Images underwent intensity inhomogeneity correction and standardization. T1w images were registered linearly to the MNI152 template. FLAIR images were linearly mapped to T1w in MNI space.

Classifier design. Volumetric datasets served as inputs to a two-stage cascaded CNN (Figure 1): the first CNN was designed to maximize sensitivity (*i.e.*, detecting a maximum number of lesional voxels), while the second optimized specificity (*i.e.*, reducing false positives).

Validation. A 5-fold cross validation repeated 5 times tested sensitivity (prediction co-localizing with manual FCD labels). Specificity was assessed by testing the model on 38 healthy controls (age: 30 ± 7) and 63 temporal lobe epilepsy (TLE) patients as disease controls (age: 31 ± 8), matched for age and sex to the training cohort.

RESULTS

The 5-fold cross-validation of the Bayesian CNN classifiers resulted in a sensitivity of 89%, with an average of 50/56 lesions detected, compared to 82% using the non-Bayesian CNN, at an identical cluster-wise false positive rate. Non-parametric permutation tests (one-tailed, 10,000 iterations) assessing the pair-wise predictive accuracy based on area under the curves (AUCs) showed that sensitivity of the Bayesian CNNs was significantly higher than the non-Bayesian (see Table 1).

CNN classifier	Sensitivity	FP	Dice	AUC Permutation Tests
Non-Bayesian (C1)	82% (46/56)	4±5	0.49	—
Bayesian (C2; mean-based threshold)	89% (50/56)	5±4	0.47	C2 > C1 ($p < 0.05$)
Bayesian (C2; uncertainty-based threshold)	89% (50/56)	5±5	0.47	C3 > C1 ($p < 0.05$)

Table 1. Performance metrics for the three CNN classifiers. Sensitivity is derived after averaging across 5 trials and thresholding to aggregate voxel as clusters. The rate of false positives (FP) clusters is averaged across patients. The Dice index represents average FCD lesion coverage compared to manual labeling.

Voxel-wise receiver operating characteristic (ROC) curves are shown in Figure 2A. Higher AUC scores signify better classification performance. Uncertainty values positively correlated with predictive probabilities at the individual level for both the mean-based thresholding (healthy controls: Pearson's $r = 0.81 \pm 0.03$, $p < 0.05$; disease controls: $r = 0.77 \pm 0.04$, $p < 0.05$) and uncertainty-based thresholding (healthy controls: 0.78 ± 0.04 , $p < 0.05$; disease controls: 0.81 ± 0.03 , $p < 0.05$). Specificity in healthy controls was 84% (no findings in 32/38; 1 ± 0 FPs) for both Bayesian and the non-Bayesian CNNs; in disease controls, specificity was 87% (no findings in 55/63; 1 ± 0 FPs) with the Bayesian CNNs, and slightly higher at 92% (no findings in 58/63; 1 ± 0 FPs) using the non-Bayesian CNN.

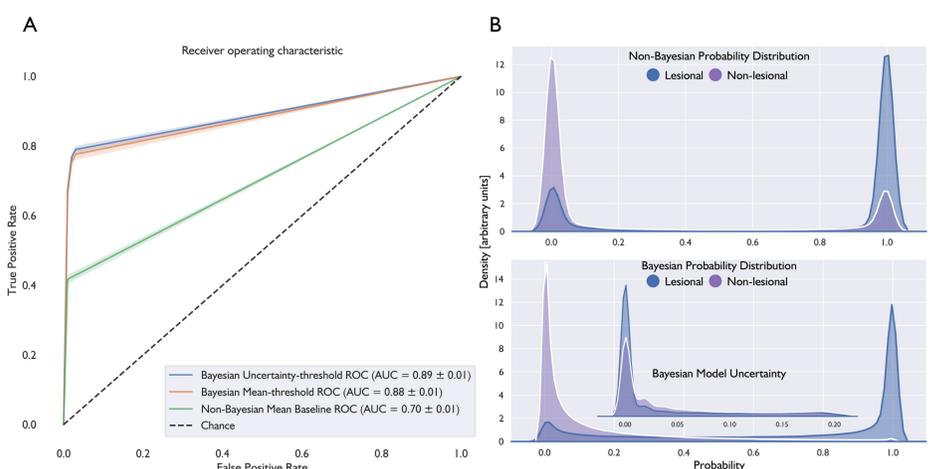


Figure 2. A. Receiver operating characteristic (ROC) curves of the three CNN classifiers. The opaque error line represents the ± 1 standard deviation of the area under the curve (AUC) around the mean AUC (solid colored line). The dotted line represents the AUC for a random classifier. B. The posterior predictive distributional profiles for FCD lesions and non-lesional tissue of the non-Bayesian CNN (top panel) and Bayesian CNN (bottom panel – only mean based thresholding depicted). The Bayesian model uncertainty is shown (inset) in the bottom panel.

CONCLUSION

We present the first deep learning method to segment FCD that leverages uncertainty for clinical decision-making with the highest sensitivity to date. Notably, epistemic uncertainty is important for safety-critical applications and instances with small datasets. Our framework exploits uncertainty both during the intermediate testing and the final prediction. Uncertainty estimates can be used to refer uncertain predictions to experts for further evaluation. This is specially important when considering that 80% of the FCD lesions detected by the CNN were missed by conventional radiological inspection. Ease of implementation, minimal pre-processing, significant performance gains coupled with uncertainty information make our CNN classifier an ideal platform for large-scale clinical use, particularly in “MRI-negative” FCD.

REFERENCES

- [1] Bernasconi A, et al. Advances in MRI for ‘cryptogenic’ epilepsies. *Nat Rev Neurol* 2011, [2] Adler S., et al. Novel surface features for automated detection of focal cortical dysplasias in paediatric epilepsy. *NeuroImage Clin* 2017, [3] Hong, SJ, et al. Automated detection of cortical dysplasia type II in MRI-negative epilepsy. *Neurology* 83, 48–55 (2014), [4] Gill RS, et al. Automated detection of epileptogenic cortical malformations using multimodal MRI. *DLMIA/ML-CDS, MICCAI Proceedings* (eds. Cardoso MJ, et al.) 349–356 (2017), [5] Tan YL, et al. Quantitative surface analysis of combined MRI and PET enhances detection of focal cortical dysplasias. *Neuroimage* 166, 10–18 (2017), [6] Kini, LG, et al. Computational analysis in epilepsy neuroimaging: A survey of features and methods. *Neuroimage: Clin* 11, 515–529 (2016). [7] Gal Y & Ghahramani Z: Bayesian Convolutional Neural Networks with Bernoulli Approximate Variational Inference. <https://arxiv.org/abs/1506.02158v6> (2015).



*ravnoor.gill@mail.mcgill.ca
<https://noel.bic.mni.mcgill.ca/>